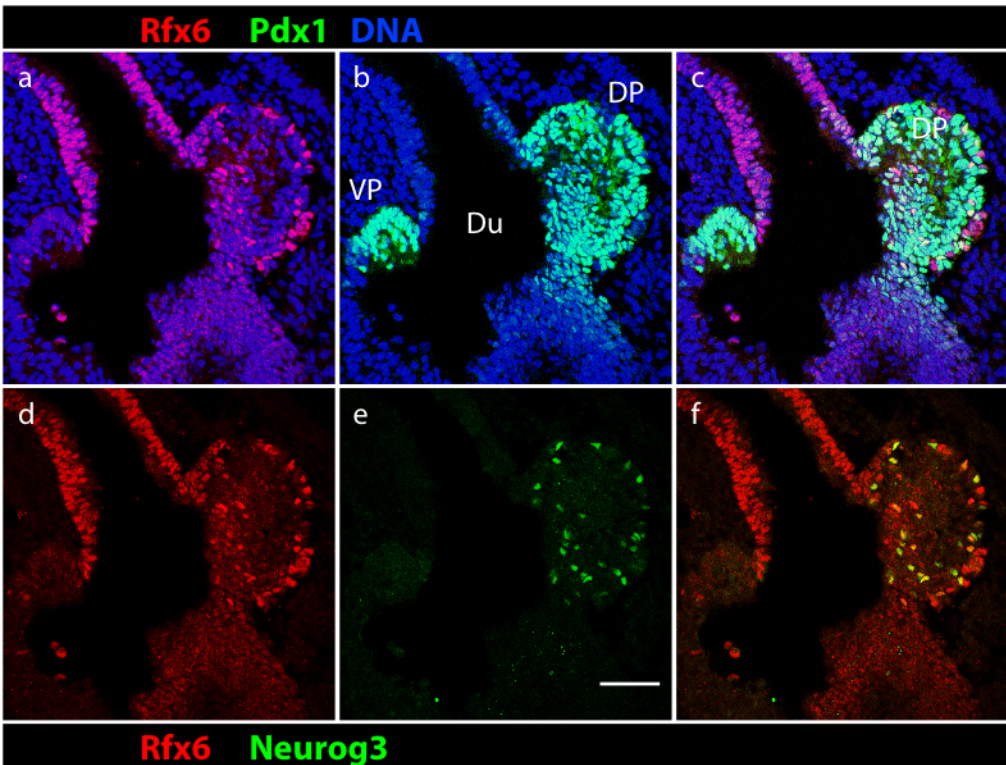
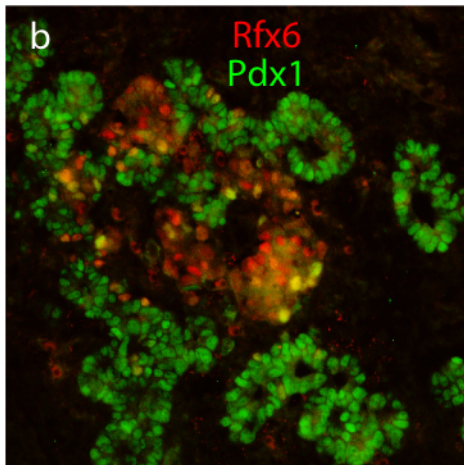
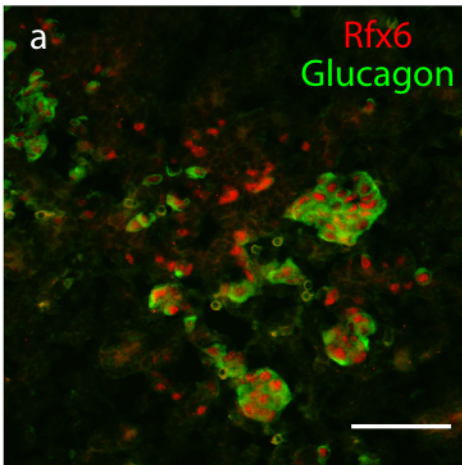


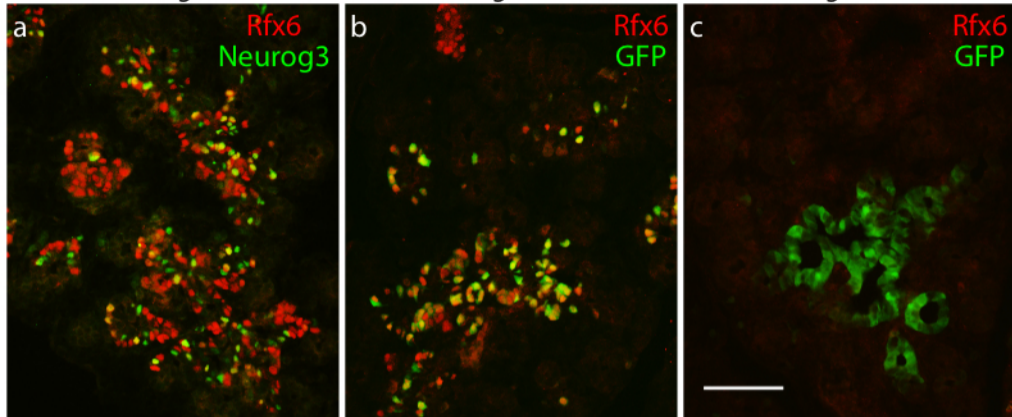
**Figure S1: Expression pattern of Rfx6 at e9.5.** Immunofluorescence staining was performed for Rfx6 (red) and Nkx2.2 (green) on a mouse embryo at e9.5. Rfx6 and Nkx2.2 staining overlaps (yellow) in the nuclei of scattered cells in the dorsal pancreatic bud (DP); Rfx6 is expressed alone in the duodenum (Du). Scale bars, 25  $\mu$ m.



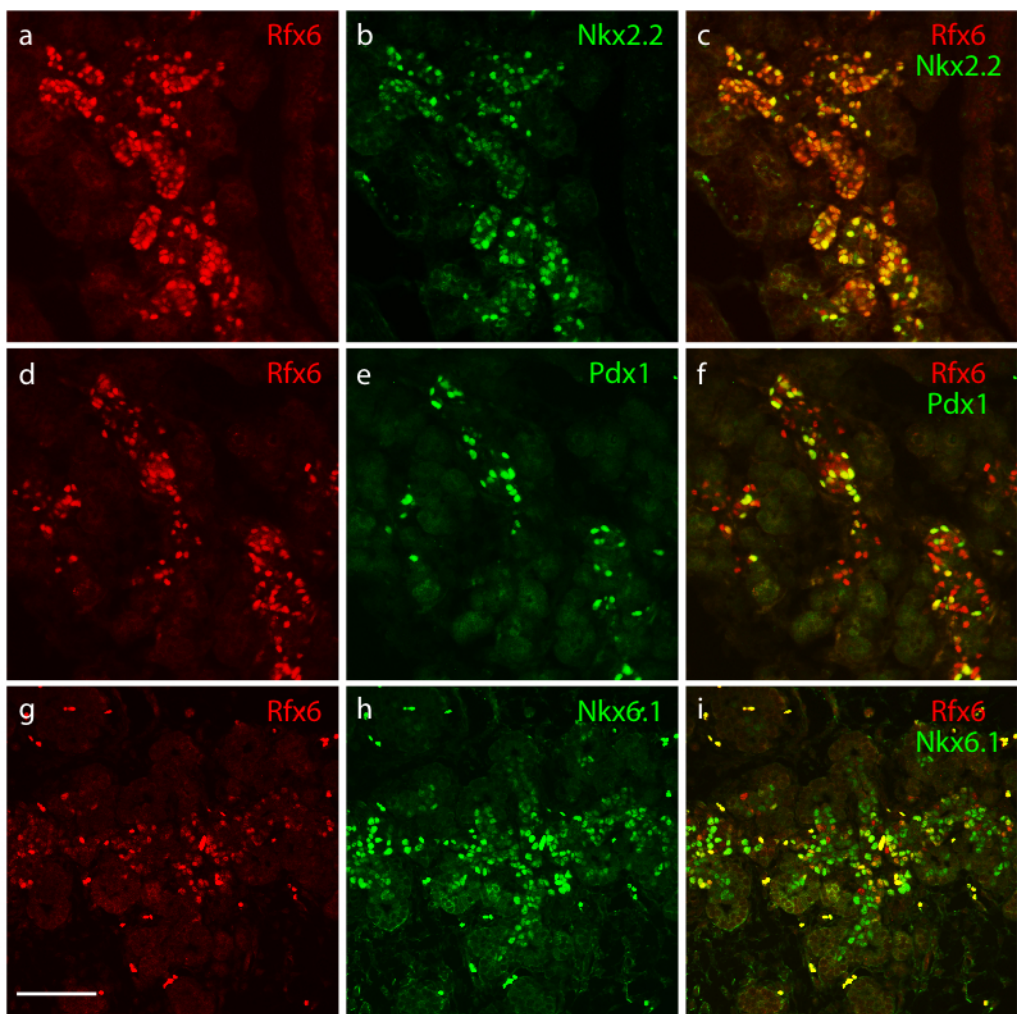
**Figure S2: Expression pattern of Rfx6 at e10.5.** Immunofluorescence staining was performed for Rfx6 (red) with Pdx1 (green, b and c) and Neurogenin3 (green, e and f) on a mouse embryo at e10.5. Nuclei are counter stained with 4',6-diamidino-2-phenylindole (DAPI). Rfx6 and Pdx1 stain the nuclei of distinct cells in the dorsal (DP) and ventral pancreatic bud (VP). Rfx6 and Pdx1 expression overlaps in the duodenum (Du) adjacent to the pancreatic buds, but Rfx6 staining extends further rostral and caudal along the gut than Pdx1. Rfx6 and Neurogenin3 staining overlaps (yellow) in the nuclei of scattered cells in the dorsal pancreas in panel f. Scale bars, 25  $\mu$ m.



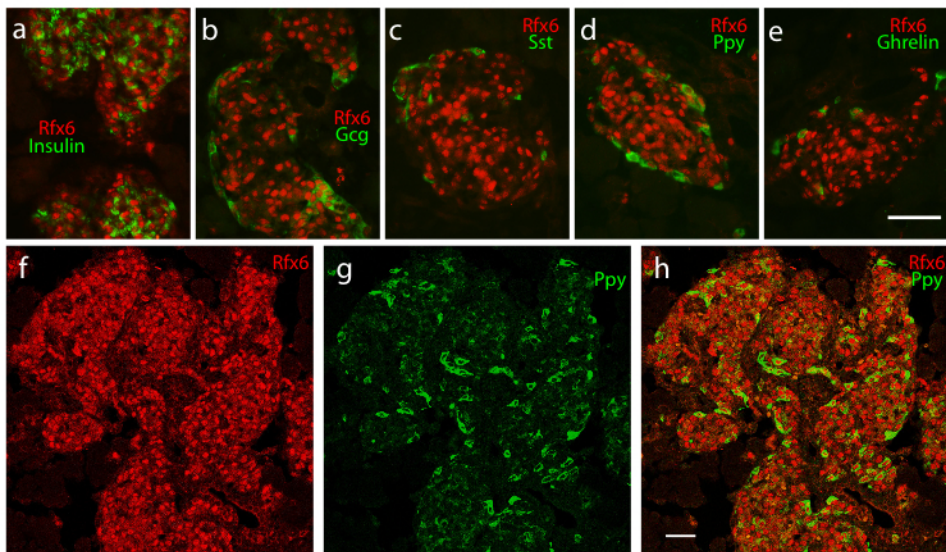
**Figure S3: Expression pattern of Rfx6 in the pancreas at e12.5.** Immunofluorescence staining was performed for Rfx6 (red) with glucagon (green, a) and Pdx1 (green, b) in the embryonic mouse pancreas at e12.5. The nuclei of the glucagon expressing cells stain for Rfx6, while Pdx1 and Rfx6 only overlap (yellow) in the nuclei of rare cells. Scale bars, 25  $\mu$ m.

*Neurog3<sup>+/+</sup>**Neurog3<sup>+/eGFP</sup>**Neurog3<sup>eGFP/eGFP</sup>*

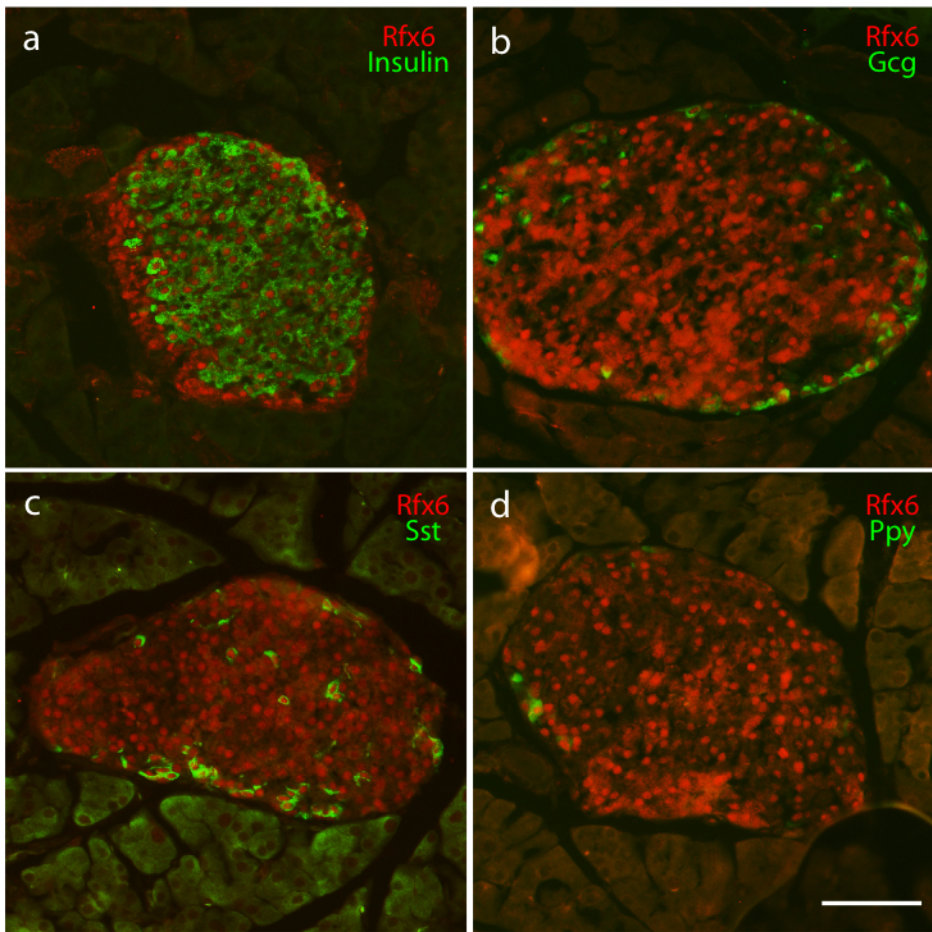
**Figure S4: Expression pattern of Rfx6 in the pancreas at e13.5.** Immunofluorescence staining was performed for Rfx6 (red) with Neurog3 (green, a) and eGFP expressed from the Neurog3 locus (green, b and c) in the embryonic pancreas harvested from *Neurog3<sup>+/+</sup>* (a), *Neurog3<sup>+/eGFP</sup>* (b), and *Neurog3<sup>eGFP/eGFP</sup>* (c) mice<sup>9</sup> at e13.5. Rfx6 nuclear staining overlaps with Neurog3 in the *Neurog3<sup>+/+</sup>* pancreas (a) and with eGFP in the *Neurog3<sup>+/eGFP</sup>* pancreas (b), but is not detectable in the *Neurog3<sup>eGFP/eGFP</sup>* pancreas. Scale bars, 25  $\mu$ m.



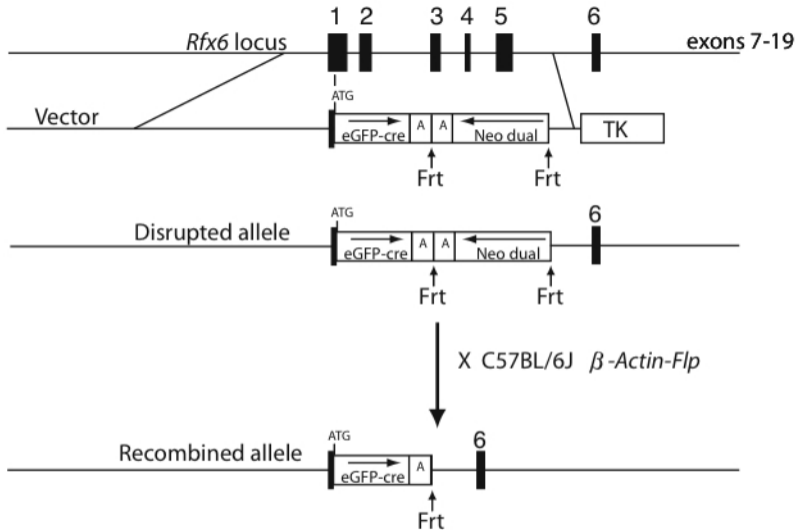
**Figure S5: Expression pattern of Rfx6 in the pancreas at e15.5.** Immunofluorescence staining was performed for Rfx6 (red) with Nkx2.2 (green, b and c), Pdx1 (green, e and f) and Nkx6.1 (green, h and i) in the embryonic mouse pancreas at e15.5. Scale bars, 25  $\mu$ m.



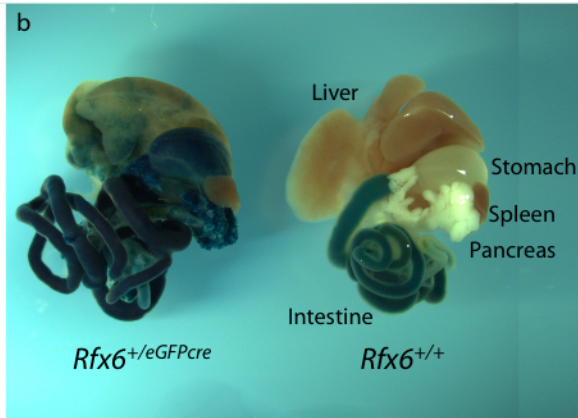
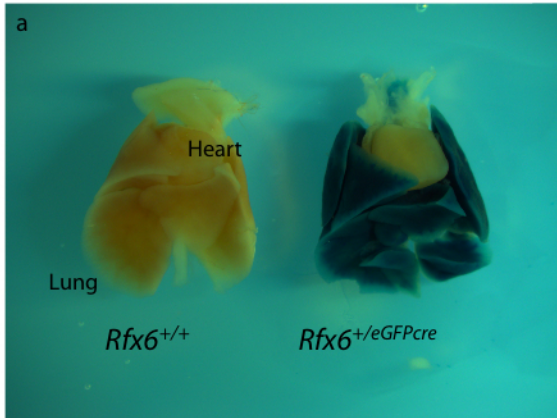
**Figure S6: Expression pattern of Rfx6 in the pancreas at e18.5.** Immunofluorescence staining was performed for Rfx6 (red) with insulin (green, a), glucagon (Gcg, green, b), somatostatin (Sst, green, c), pancreatic polypeptide (Ppy, green, d) and ghrelin (green, e) in the embryonic mouse pancreas at e18.5. In panels f - g, confocal imaging was used to localize accurately Rfx6 (red) in the nuclei of cells expressing pancreatic polypeptide (Ppy, green). Separate colour channels are shown for red (f) and green (g). Scale bars, 25  $\mu\text{m}$ .



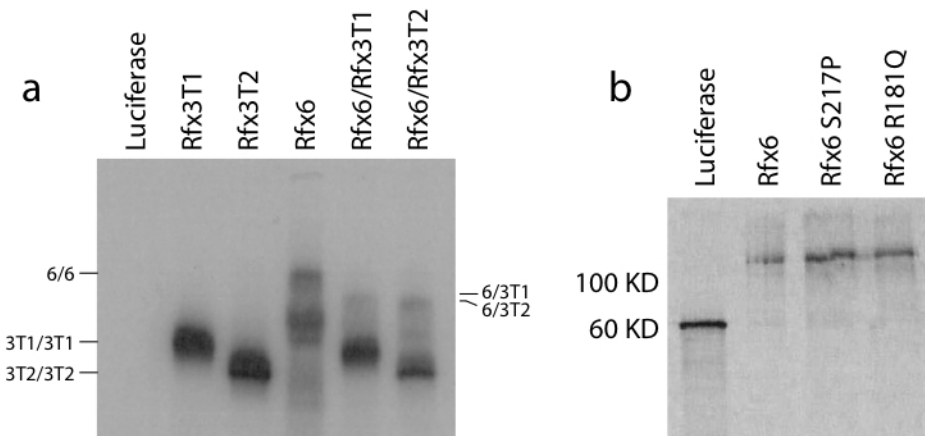
**Figure S7: Expression pattern of Rfx6 in the adult pancreas.** Immunofluorescence staining was performed for Rfx6 (red) with insulin (green, a), glucagon (Gcg, green, b), somatostatin (Sst, green, c) and pancreatic polypeptide (Ppy, green, d) in the adult mouse pancreas. Scale bars, 25  $\mu$ m.



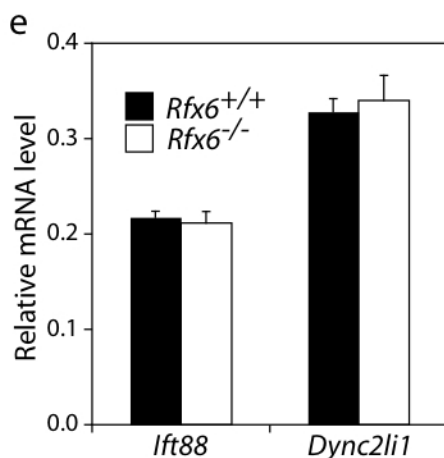
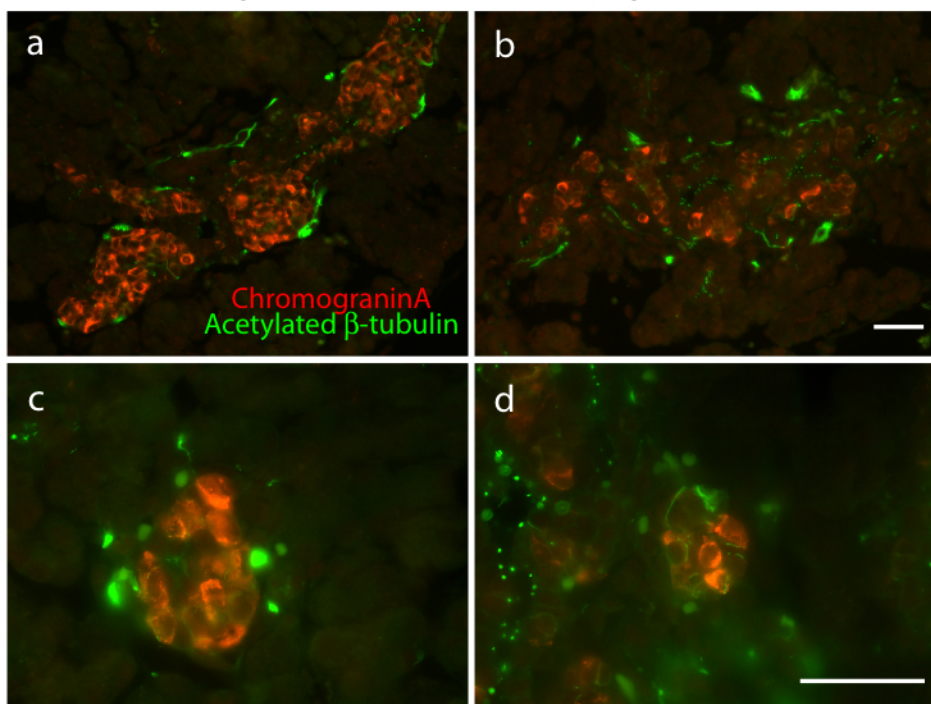
**Figure S8: Targeting of the *Rfx6* locus in mice.** The design is shown for the *Rfx6*<sup>eGFPcre</sup> targeting construct. The first 5 exons were replaced with a marker gene encoding an eGFPcre fusion protein.



**Figure S9: Lineage tracing of the *Rfx6*-expressing cells.** Lineage tracing was performed on *Rfx6*<sup>+/+</sup>/R26R (right in a and left in b) on *Rfx6*<sup>+/eGFPcre</sup>/R26R (right in a and left in c) mice at p0 by staining for  $\beta$ -galactosidase activity with Xgal (blue). In a, Xgal stains the lungs, but not the heart, which derives from the mesoderm germ layer. In b, Xgal stains the entire gut, pancreas and liver, but not the mesoderm-derived spleen.



**Figure S10: Rfx3 and Rfx6 bind to DNA as heterodimer partners.** In a, DNA binding of the human *in vitro*-translated proteins shown above each lane to the double-stranded, radiolabeled oligonucleotide HBV X-box probe<sup>21</sup> was tested by electromobility shift assay (EMSA). With Rfx3 in excess, binding of Rfx6 homodimers (6/6) was completely lost and replaced with the Rfx6/Rfx3 heterodimers in lanes 4 and 5. In b, the efficiency of translation of the wildtype and mutant Rfx6 proteins was compared by including <sup>35</sup>S-labelled methionine in the translation mix and separation by gel electrophoresis of equal amounts of the translation products.

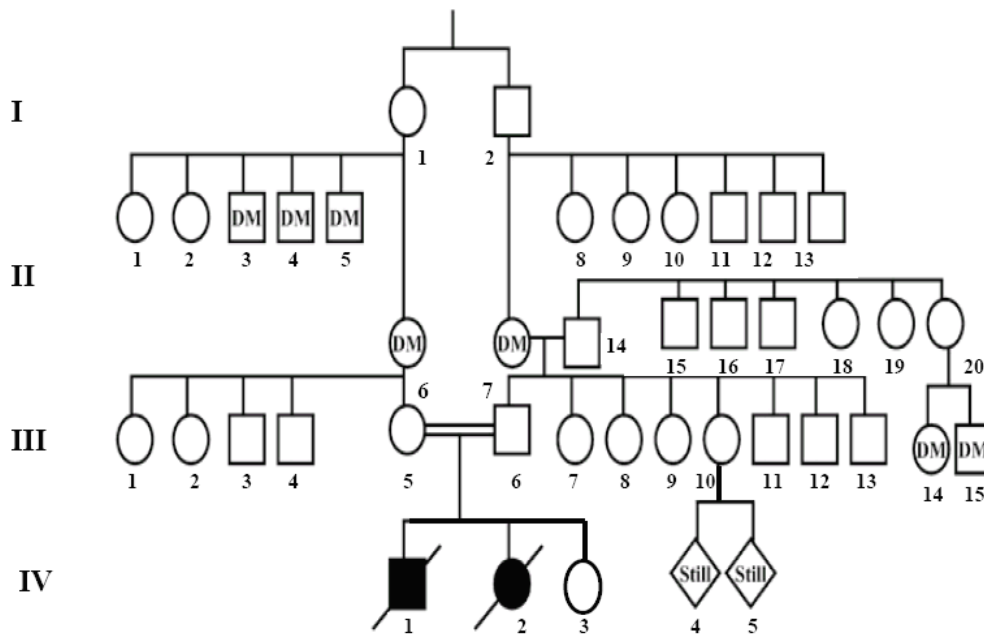
*Rfx6*<sup>+/+</sup>*Rfx6*<sup>eGFPcre/eGFPcre</sup>

**Figure S11: *Rfx6* is not required for the generation of primary cilia.** In panels a-d, immunofluorescence staining was performed for chromograninA (red) with acetylated  $\beta$ -tubulin (green), a marker for primary cilia, in pancreas harvested at e17.5. Primary cilia appear as long, fine strands linked to the chromograninA-expressing cells. Scale bars, 25  $\mu$ m. In e, the mRNA for genes expressed in primary cilia, *Ift88* and *Dync2li1* were amplified by real-time RT-PCR (TaqMan) from RNA isolated from the pancreas at e17.5 from mouse embryos with the genotypes shown and expressed relative to the level of *Gusb*.

## ON-LINE SUPPLEMENTARY MATERIAL FOR THE HUMAN MAPPING

### Summary of clinical cases:

**Proband # 1:** Case described in Mitchell et al.<sup>17</sup> IV-2 in this pedigree. A baby girl born to Pakistani second cousins at 34 weeks of age after induced delivery due to poor fetal growth (<3<sup>rd</sup> percentile). She was suspected by prenatal ultrasound to have duodenal atresia, which was confirmed by surgery at day 1 of life along with jejunal atresia, annular pancreas and gall bladder agenesis. Neonatal diabetes was diagnosed at 39 hours of life with hyperglycemia, low insulin and C-peptide levels. In addition to insulin therapy, she was on TPN (total parental nutrition) due to malabsorption and intractable diarrhea not responsive to pancreatic enzyme supplementation. Her course was further complicated by liver failure. A liver biopsy revealed cholestasis consistent with TPN and/or extra-hepatic biliary obstruction. She died at 194 days due to multi-organ failure. Parents approved an autopsy. Histologically, the exocrine pancreas appeared normal. Immunohistochemical staining revealed occasional chromogranin A positive cells in the absence of any well-formed islets or insulin, glucagon or somatostatin positive staining cells. The family history is relevant for a brother (case 1 in Mitchell et al.<sup>17</sup>) who had the same diagnosis and course. No usable DNA could be extracted from intestinal tissue in paraffin blocks from the atresia surgery of that sibling, the only material available from him. Several cases of type 2 diabetes, two in obligate carriers but the parents were not tested (Fig S12)



**Figure S12.** IV-2 is proband #1 in our study, parents are second cousins. In this large family, there were no other known affected children. Pedigree taken from Mitchell et al.<sup>17</sup> with minimal changes. DM=diabetes mellitus, Still = stillborn.

**Proband # 2:** This case has been reported previously by Chappell et al.<sup>18</sup>. A baby girl born small for gestational age (2<sup>nd</sup> percentile) at 35 weeks to first-degree Pakistani cousins. Duodenal atresia was surgically corrected at day 1 of life. Gall bladder agenesis was also noted during surgery. Insulin treatment was begun at day 8 of life after persistent hyperglycemia with low C-peptide levels. Despite normal feeding, she failed to thrive and developed hyperbilirubinemia. Liver biopsy showed

cholestatic disease while no biliary atresia was noted on cholangiography but the portal areas were oedematous and narrowed by bile plugs. Gradually, the hyperbilirubinemia resolved without any sign of liver disease by 11 months of age. She is still alive at 4 ½ years and insulin dependent. No relevant family history reported. .

Proband # 3: This is case 5 in Mitchell et al<sup>17</sup>. A baby girl, product of in vitro fertilization with a donated egg, born small for gestational age at 39 weeks. Duodenal atresia was suspected on prenatal ultrasound. At day 2 of life, she underwent laparoscopy revealing malrotation and duodenal web necessitating lysis of bands. During this hospitalization she developed hyperglycemia and received insulin briefly. She was readmitted at 1 ½ months with hyperbilirubinemia, failure to thrive and hyperglycemia requiring insulin with low C-peptide levels. Liver biopsy was consistent with cholestatic disease with normal intra- and extra hepatic ducts. She is alive at 9 years and insulin dependent but on a small dose. No family history available.

Proband # 4: This case has not been previously published. A baby boy born small for gestational age (<10<sup>th</sup> percentile) at 35 weeks for non-consanguineous French parents. Prenatal ultrasound revealed poor fetal growth and duodenal atresia later confirmed by surgery performed at day 9 of life. Neonatal diabetes was diagnosed at day 2 in the setting of hyperglycemia with undetectable insulin and C-peptide levels. His course was complicated by refractory ascites, sepsis and gastrointestinal hemorrhage, of which he died at 2 ½ months of age. The family history is relevant for gestational diabetes in the mother and diabetes in the father, diagnosed as type 1.

Proband #5: (Désir et al., case report submitted). Male born in Belgium of parents of Moroccan descent with intestinal atresia, intestinal malrotation, intrauterine Growth Restriction (IUGR), gall bladder agenesis, and hemochromatosis. Treatment with insulin began from the third day of life. C peptide levels were low. The child died at three months of age.

Proband # 6: Not previously published. Female patient born to a non-consanguineous Irish parents. Increased amniotic fluid and intrauterine Growth Restriction (IUGR) were noticed during pregnancy. She was born with duodenal atresia. After the atresia repair operation, neonatal diabetes was diagnosed and required insulin treatment.

### **Design of Nimblegen array:**

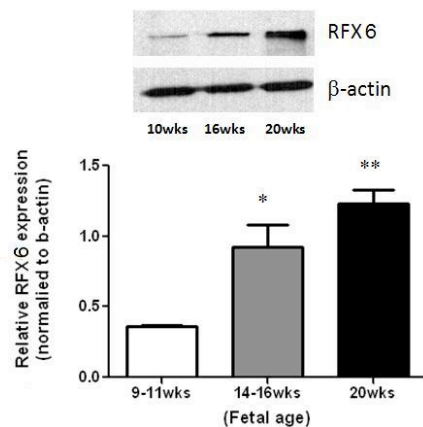
We identified 1,322 exonic sequences mapping to the positionally ascertained loci, using the UCSC Genes track (a compilation of RefSeq, Genbank, CCDS, and UniProt) of the hg18 March 2006 human genome assembly and annotation

Oligonucleotide probes (> 60 bases) were designed to tile the exonic sequences plus 50 bp on either side. The Nimblegen algorithm excludes any probe containing a 15-mer that occurs more than 100 times in the human genome. In addition, probe sequences containing close matches elsewhere in the genome are removed using the SSAHA algorithm (<http://www.sanger.ac.uk/Software/analysis/SSAHA/>).

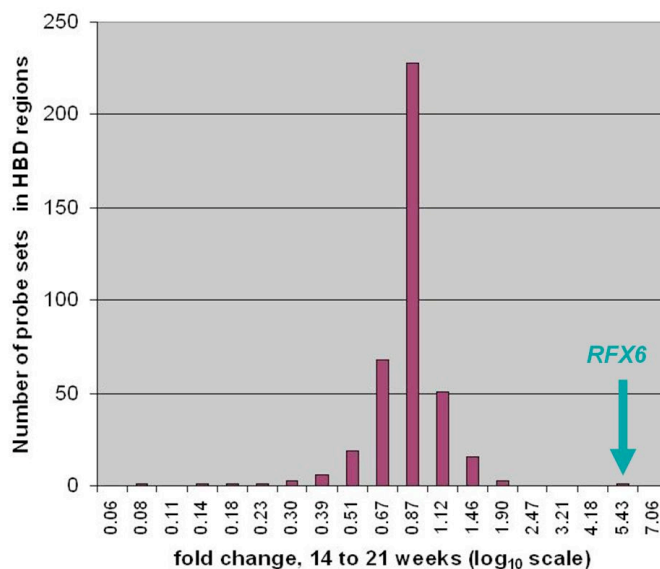
Further details are provided in the Nimblegen website:

[http://www.nimblegen.com/products/lit/probe\\_design\\_2008\\_06\\_04.pdf](http://www.nimblegen.com/products/lit/probe_design_2008_06_04.pdf)

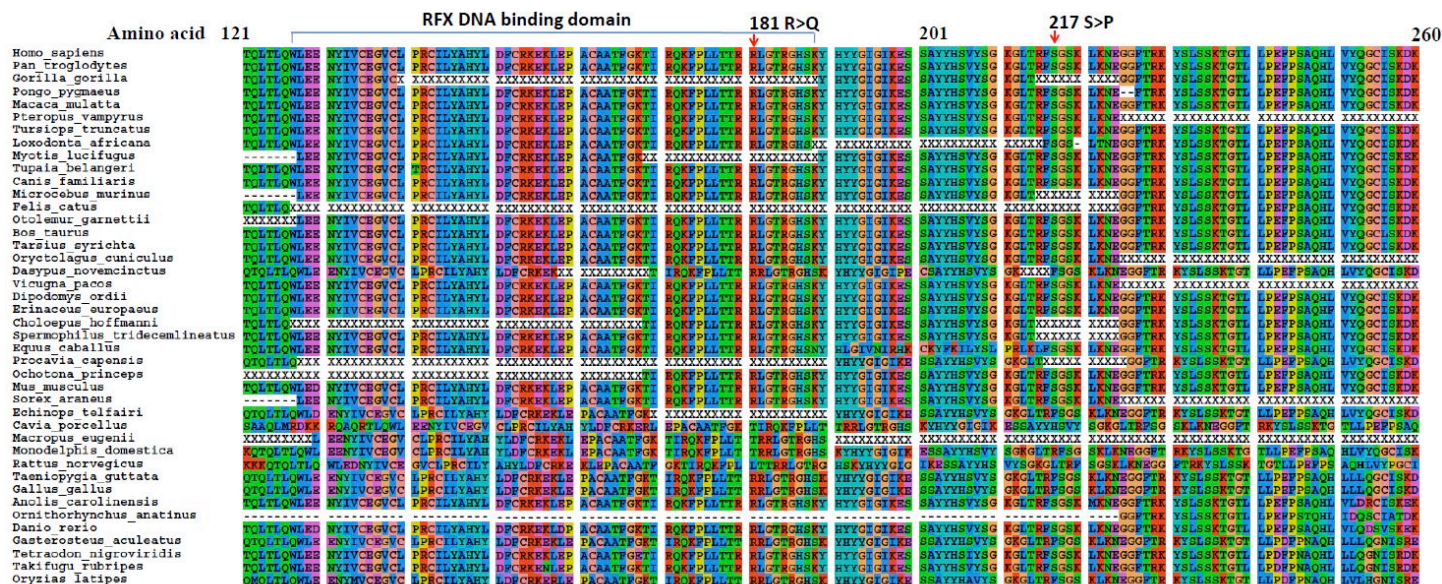
The user has the option of overriding these non-uniqueness filters. We did not, as only 36 tiles were excluded, leaving 1,262 tiles encompassing 1,309 exons. Whole-genome amplified DNA from Proband #2 was fragmented and hybridized to the array and the eluate was amplified and sequenced by the Roche 454 process.



**Figure S13.** Total protein from human fetal pancreatic tissues was extracted by sonication in Nonidet-P40 lysis buffer. Equal amounts (30 $\mu$ g) from different development ages were separated by 10% SDS-PAGE electrophoresis and immunoblotted on nitrocellulose membranes with mouse anti-human RFX6 antibody and goat anti-mouse horse radish peroxidase-conjugated secondary antibody. Densitometric quantification of bands at subsaturation levels was performed using Syngenetool gel analysis software (Syngene, Cambridge, UK) and normalized to  $\beta$ -actin as a loading control, with at least three replicates per age group.



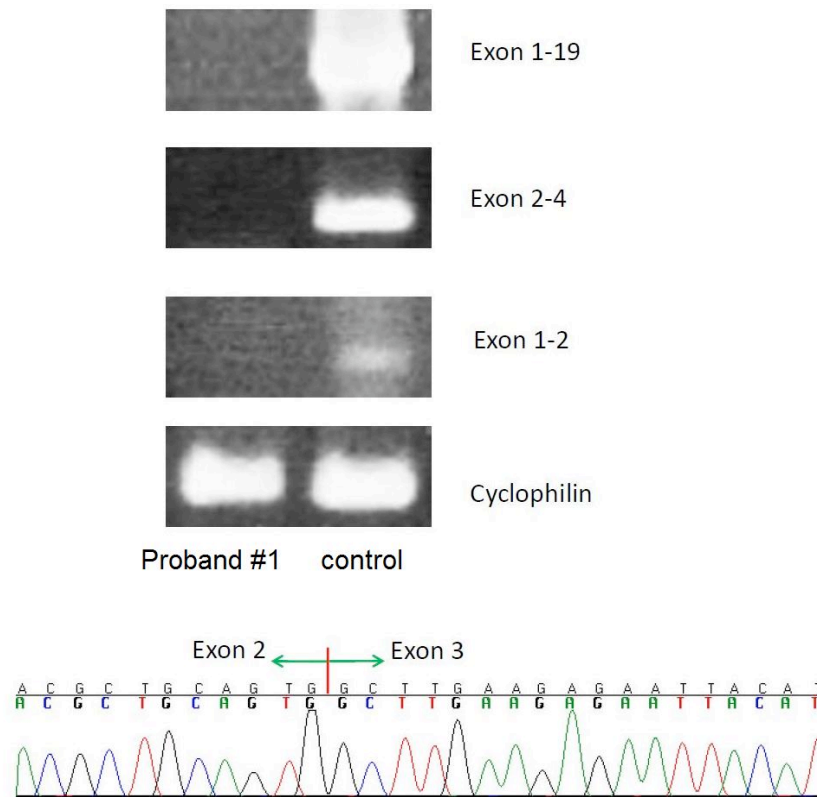
**Fig. S14.** Frequency histogram of fold-change in the expression levels of the genes mapping to the overlapping HBD regions in human fetal pancreas between 14 and 21 weeks (R. Wang, unpublished microarray data). The X-axis labels refer to the lower end of the interval.



**Figure S15**

Multiple sequence alignments of orthologues from vertebrate species from the Ensembl database (<http://www.ensembl.org>). Graphics by Seaview software, Galtier et al., *Comput Appl Biosci* 12(6): 543-8 1996 (<http://pbil.univ-lyon1.fr/software/seaview.html>). The protein sequence is highly conserved between different species. The genomic evolutionary rate profiling (GERP) (Cooper et al. *Genome Research* 15(7): 901-913 2005) conservation score is 1.00 for the 217 position and 0.965 for the entire *RFX6* DNA binding domain, while the 181 position has a score of 0.838.

The DNA binding domain, where 181R is located, is also highly conserved within the human *RFX* family and 181R is predicted to be in direct contact with the DNA (Aftab et al., 2008, *BMC Evol Biol* 8:226, doi:1471-2148-8-226).



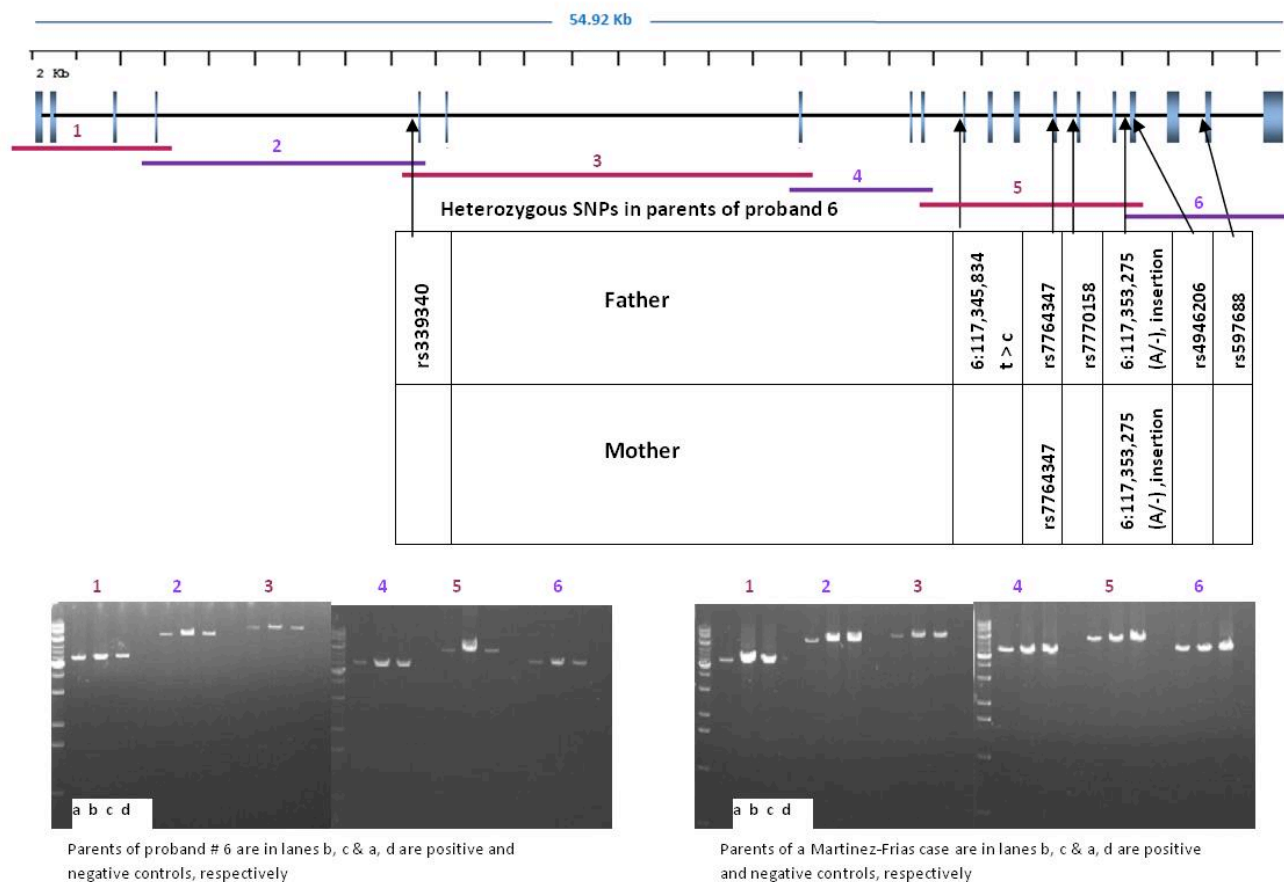
### Figure S16

RNA extracted from frozen autopsy pancreas of Proband #1 was used to evaluate the effect of the homozygous IVS2+2 splicing mutation. First, the entire *RFX6* cDNA, exons 1-19 was easily amplified from normal human foetal pancreas (control). Fragment size (3,418 nt) and sequencing of the only transcript amplified confirmed all predicted splicing sites (NM\_173560: 241). The normal sequence of the junction of exons 2-3 (affected by the mutation) is shown at the bottom of the figure, for a normal pancreas.

Smaller fragments, encompassing exons 2-4 (expected to be altered by the mutation) and 1-2 (should not be affected) were also amplified in normal pancreas. No detectable amplification was seen in pancreas RNA of Proband #1 with any of the primer pairs, consistent with nonsense-mediated decay. Quality and loading of the RNA was confirmed by robust amplification of cyclophilin.

#### Primers:

Exons 1- 19 (3,178 bp)	forward 5'- GGGGAGAAAGGCGAAGAC-3'	reverse 5'- CGGAATGCTGCTATGTGAC-3'.
Exons 2- 4 (242 bp)	forward 5'- ACCATCACGCAGATTGTGAA -3'	reverse 5'- TTTGAATGGCCTCTTGTTC -3'.
Exons 1-2 (336 bp)	forward 5'- AAGACTGCTGTGTGCAGCTC-3'	reverse 5'- CAGGTATTGATCCGCTGCTT-3'.



**Figure S17.** Long-range PCR showing absence of large deletions in the parents of proband #6 and a case of the Martinez-Frias syndrome. The indicated overlapping fragments were amplified from genomic DNA and cover the entire gene. The presence of two different heterozygous deletions in each parent is made unlikely by the multiple heterozygosities in the intronic segments sequenced in the mutation search